Structural studies of bacterial steroid-hydroxylating cytochrome P450 monooxygenases

Jóźwik, Ilona

Chapter 4

Crystal structure of the vitamin D_3 hydroxylating enzyme CYP109A2

Ilona K. Jóźwik[1], Flora M. Kiss[2], Bauke W. Dijkstra[1], Elisa Brill[2], Rita Bernhardt[2] and Andy- Mark W. H. Thunnissen[1]

[1] Laboratory of Biophysical Chemistry, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

[2] Department of Biochemistry, Saarland University, Saarbrücken, Germany

In preparation for publication

CYP109A2 from *Bacillus megaterium* DSM319 is a cytochrome P450 monooxygenase able to convert vitamin D3 (VD3) to 25-hydroxyvitamin D3 as a single product. It has no activity on steroid substrates like corticosterone or testosterone. In contrast, the closely related CYP109E1 enzyme from the same organism (46% sequence identity) hydroxylates VD3 at positions 24 and 25, and can convert testosterone to 16β-hydroxytestosterone. To explain the remarkable differences in substrate selectivity and regio-specificity of CYP109A2 and CYP109E1 we studied these enzymes using protein X-ray crystallography, molecular modeling and biochemical methods. Previously, we reported the crystal structures of CYP109E1 in a steroid-free state, and with bound corticosterone and testosterone. Here we report three crystal structures of CYP109A2: one of its substrate-free form determined at 2.7 Å resolution and two resulting from crystallization in the presence of VD3, determined at 2.2 Å and 1.77 Å resolution. All CYP109A2 structures crystallized in an open conformation and display an overall fold highly similar to the substrate-free structure of CYP109E1. Compared to CYP109E1, CYP109A2 contains a shorter and differently folded BC loop, its active site is more polar, and access of substrate to the heme cofactor is more restricted. Although extra electron density is visible in the active sites of the 2.2 and 1.77 Å CYP109A2 structures, it was not possible to confirm the presence of bound VD3. A comparison of CYP109A2 and CYP109E1 to other bacterial P450s converting VD3 further indicates that they are closely related to CYP107BR1 (Vdh) from *Pseudonocardia autotrophica* with respect to amino acid sequence, volume of the active site pocket and location of VD3 binding residues. The folds and sequences of the BC loops in these three P450s are highly divergent and may account for the differences in substrate specificity.

Introduction

Cytochromes P450 (P450s) are highly versatile heme-thiolate catalysts that oxidize various organic compounds by selective activation of non-reactive C-H bonds, followed by the insertion of one atom of molecular oxygen while reducing the other oxygen atom to water. These enzymes are present in nearly all organisms where they catalyze essential reactions related to steroid hormone metabolism and xenobiotic biotransformations [1]. Since P450s generally show a broad substrate and reaction specificity, they are considered as powerful 'green' tools for the synthesis of pharmaceutically and biotechnologically desired compounds, offering an alternative to difficult and environmentally unfriendly chemo-synthetic routes. For such applications, microbial cytochromes P450 are of special interest, as they are generally more soluble and stable than their eukaryotic counterparts. In particular, the potential of microbial P450s to produce pharmaceutically important steroid derivatives is being explored intensively [2-4].

Chemical synthesis of steroidal compounds in a regio/stereo-selective manner is challenging and usually requires several steps, limiting cost-effectiveness and final yields. One example is the chemical conversion of cholesterol to functional derivatives of VD_3, such as 25-hydroxyvitamin D_3 ($25(OH)VD_3$) or 1α,25-dihydroxyvitamin D_3 ($1\alpha,25(OH)_2VD_3$), which involves approximately 20 steps and is highly inefficient and expensive, with total yields not exceeding 1% [5,6]. Such obstacles may be circumvented by application of specific bacterial P450s oxidizing VD_3 to $25(OH)VD_3$ or $1\alpha,25(OH)_2VD_3$, in a single enzymatic conversion step. At present, the best characterized bacterial VD_3 hydroxylases are CYP107BR1 (also known as vitamin D_3 hydrolase or Vdh) from *Pseudonocardia autotrophica* and CYP105A1 (also known as $P450_{SU-1}$) from *Streptomyces griseolus* [7]. To improve their activity and regio-selectivity, these enzymes have been subject to extensive protein engineering [8-11]. Crystal structures are available of both enzymes, in substrate-free and substrate-bound states. For Vdh, a quadruple mutant obtained by directed evolution, called Vdh-K1 (T70R/V156L/E216M/E384R), was crystallized with bound VD_3 or $25(OH)VD_3$. The substrates bind productively near the heme-iron, with an extended conformation and opposite orientation, in accordance with hydroxylations at position 25 for VD_3 and position 1α for $25(OH)VD_3$. For CYP105A1, the product $1\alpha,25(OH)_2VD_3$ was captured in the crystal structures of two engineered mutants, but its distance from the heme-iron is too far to explain the observed regioselectivity. Structural work on CYP105A1 supported by site-directed mutagenesis identified three arginine residues in the binding pocket that influence VD_3 recognition – R73 and R84 (BC loop), and R193 (G helix). R193 was found to be important for hydroxylation reactions, whereas mutation of the other two into small hydrophobic residues like alanine or valine significantly improved the catalytic efficiency of CYP105A1 and enhanced both 1α- and 25-hydroxylation activities.

Table 1. Substrates and products of the reactions catalyzed by CYP109E1 and CYP109A2 from *B. megaterium* DSM319 [12-14]. Products were identified by NMR.

Substrate	Products	
	CYP109E1	**CYP109A2**
VD_3	$25(OH)VD_3$, $24S(OH)VD_3$, $24S,25(OH)_2VD_3$	$25(OH)VD_3$
Testosterone	16β-(OH)-testosterone, androstenedione	not converted

Recently, two novel microbial VD_3 hydroxylases were identified and characterized, the P450s CYP109E1 and CYP109A2 from *Bacillus megaterium* DSM319 [12]. Initially, CYP109E1 was reported to convert testosterone to 16β-hydroxytestosterone and crystal structures were determined of the substrate-free enzyme and of complexes with bound testosterone and corticosterone [13]. More recently, CYP109E1 was shown to also catalyze the conversion of VD_3 to 7 products, including $25(OH)VD_3$, 24S-hydroxyvitamin D_3 ($24S(OH)VD_3$) and 24S,25-dihydroxyvitamin D_3 ($24S,25(OH)_2VD_3$) as major products (Table 1). Docking calculations supported by site-directed mutagenesis led to the identification of a few active site residues determining the regioselectivity of VD_3 conversion by CYP109E1 [14]. In contrast, little is known about the structure and function of CYP109A2. This latter enzyme is able to hydroxylate VD_3 to a single, mono-hydroxylated product, *i.e.* $25(OH)VD_3$, while preliminary analysis indicates no activity towards testosterone (Table 1) [12]. The aim of the present study is to increase the functional and structural knowledge of CYP109A2 by better defining its substrate scope, by solving its crystal structure(s) and by comparing its structure-function relationships to those of CYP109E1 and other relevant bacterial P450s.

Results

Sequence homology of CYP109A2 with CYP109E1 and other vitamin D_3 hydroxylating P450s

The *cyp109a2* gene was recently identified in *B. megaterium* DSM319, together with three other genes encoding cytochromes P450 (CYP102A1, CYP106A1 and CYP109E1) [12,13,26]. The sequence of CYP109A2 contains three characteristic sequence motifs (highlighted in Figure 1), which support its annotation as a P450 enzyme: (i) the 'heme-binding domain' signature just before the L helix (the 'FxxGxHxCxG' – signature sequence of bacterial P450s, helix labeling according to standard P450 conventions [27]), (ii) the 'ExxR' motif in the K helix, and (iii) the highly conserved 'ET' motif or 'acid-alcohol' pair (Glu244, Thr245) in the I helix. CYP109A2 shares the highest sequence identity with CYP109E1 from the same organism (45% identity, Figure 1). The former enzyme has also significant sequence homology with other bacterial P450s hydroxylating VD_3 or its derivative(s) *i.e.* CYP107BR1 from *P. autotrophica* (37% identity), CYP107CB2 from *B. lehensis* G1 (36% identity), CYP105A1 from *Streptomyces griseolus* (32% identity) and CYP105A2 from *P. autotrophica* (30% identity). As expected, CYP109A2 is less closely related to eukaryotic P450s that are natural VD_3 hydroxylases, sharing, for example, only 25% sequence identity with human CYP2R1.

In CYP109A2, the highest degree of conservation is shown by residues with a role in heme binding and those belonging to the substrate recognition site 4 (SRS4) (central part of I helix), where residues have roles in substrate binding, oxygen activation and proton delivery during catalysis [28]. The remaining regions associated with substrate binding (SRS2, SRS5 and SRS6) are less conserved, although relatively higher conservation levels may be observed in SRS5 or SRS6 upon pairwise comparison of P450s belonging to the same CYP family, like CYP109A2 and CYP109E1. Nevertheless, the highest degree of variability is displayed by the amino acids belonging to SRS1 (the BC loop region) and SRS3 (the G helix) (Figure 1).

Summarizing, the most conserved regions in CYP109A2 are linked to heme binding, redox partner binding or dioxygen activation and proton delivery, while the elements involved in substrate binding and specificity are less conserved.

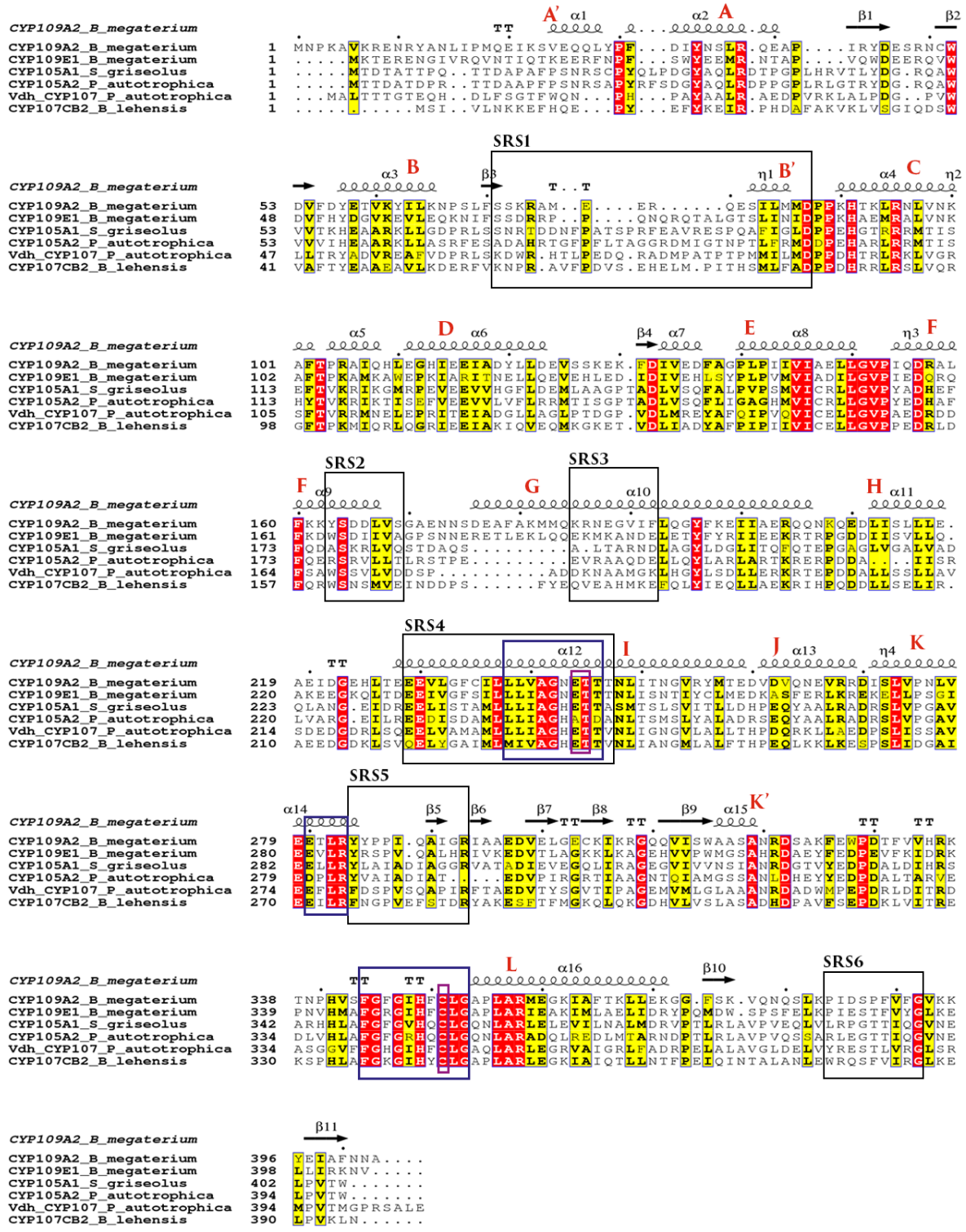

Figure 1. Multiple sequence alignment of CYP109A2 and CYP109E1 from *B. megaterium* DSM319 with other known bacterial cytochrome P450 VD_3-hydroxylases. Secondary structural elements are shown as in substrate-free CYP109A2. Conserved and similar residues are highlighted in *red* and *yellow*, respectively. Helices are labeled A-L. Residues in the six substrate recognition sites (SRS) are enclosed in *black* boxes. The characteristic P450 sequence motifs are in *blue* boxes: heme-binding domain signature just before the L helix, central part of I helix with the highly conserved 'acid-alcohol pair' of residues, and the 'ExxR' motif in the K helix.

Absence of steroid converting activity for CYP109A2

Since two of the recently identified P450s from *B. megaterium* DSM319, CYP109E1 and CYP106A1, are able to convert steroid compounds [13,29,30], we analyzed steroid conversion by CYP109A2. A library of 16 steroid compounds was screened for binding to CYP109A2 using the standard P450 spectral shift assay (Table 2). None of the tested compounds induced a type I spectral change (Soret absorption band increase at 390 nm and concomitant decrease at 420 nm), which indicates that none of the steroids is able to displace the heme-iron coordinated water molecule present in the resting state of the enzyme. Displacement of this water molecule by a substrate, and the concurrent change from a low-spin to high-spin state of the ferric iron, is considered to be a crucial step in the catalytic cycle of P450s. Instead, three tested steroids (corticosterone (**2**), deoxycorticosterone (**3**) and dexamethasone (**12**)) induced a reverse type I spectral change upon titration (shift from a high-spin state at ~390 nm to a low spin state at ~420 nm), suggesting that these ligands might adopt unproductive or inhibitory binding modes, or that possibly an endogenous type I substrate is being displaced. Since the spectral shift assay is not always a reliable method to identify substrates of P450s, and VD_3 does not induce a typical type I spectral change with CYP109A2 [12,31], we decided to test a focused library of 8 compounds (androstenedione (**1**), corticosterone (**2**), deoxycorticosterone (**3**), β-estradiol (**4**), estrone (**5**), testosterone (**6**), testosterone acetate (**7**) and 17α-methyltestosterone (**8**)) in an *in vitro* conversion assay with CYP109A2. Reconstitution of the P450 activity was accomplished using bovine adrenodoxin and adrenodoxin reductase ($Adx_{4\text{-}108}$ and AdR) as surrogate redox partner proteins, which were previously shown to support the conversions by CYP106A1 and CYP109E1 [13,29]. Moreover, VD_3 conversion was successfully established with CYP109A2 and the $Adx_{4\text{-}108}$ /AdR redox pair, serving as a positive control (data not shown). None of the tested steroids were converted by CYP109A2 with a significant yield (>3%). Hence, CYP109A2, although sharing relatively high sequence similarity to other steroid-hydroxylases, is not able to convert steroids.

Table 2. Steroids tested in this study as putative substrates for CYP109A2. Sixteen steroids were screened using a spectral binding assay with CYP109A2 as described in 'Materials and methods', out of which eight were investigated further with *in vitro* conversion experiments (Rev Type I = reverse type I spectral shift; - no shift/activity observed).

	Compound	Type I shift	*In vitro* conversion
1	androstenedione	-	-
2	corticosterone	Rev Type I	- (< 1%)
3	deoxycorticosterone (DOC)	Rev Type I	-
4	β-estradiol	-	-
5	estrone	-	-
6	testosterone	-	- (< 3%)
7	testosterone acetate	-	- (< 2%)
8	17α-methyltestosterone	-	- (< 2%)
9	cortisol	-	
10	cortisone	-	
11	11-deoxycortisol (RSS)	-	
12	dexamethasone	Rev Type I	
13	prednisone	-	
14	prednisolone	-	
15	progesterone	-	
16	19-nortestosterone	-	

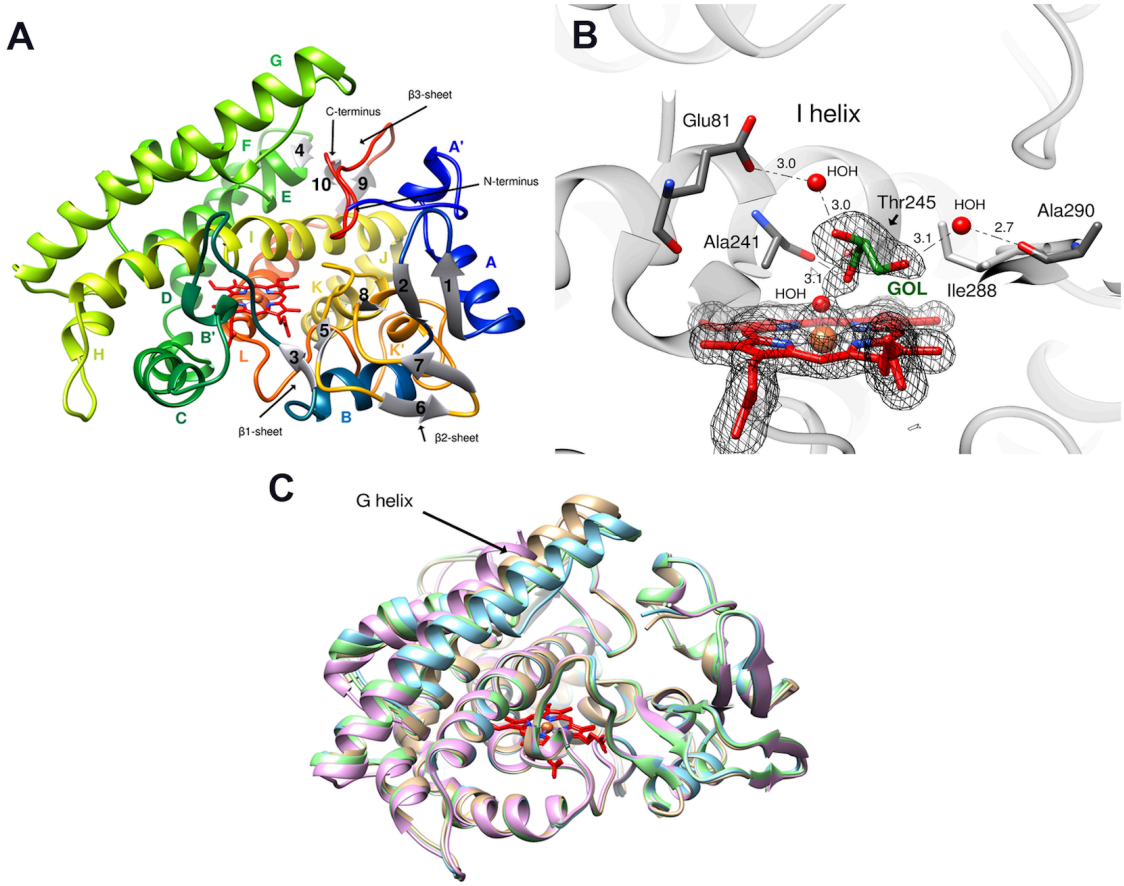

Figure 2. Structure(s) of CYP109A2 crystallized in an open conformation. A) Crystal structure of substrate-free CYP109A2 with secondary structure elements labeled following the common P450 nomenclature (rainbow coloring from *blue* N' terminus to *red* C' terminus). B) Stabilizing interactions of the bound glycerol (GOL, in *green*) in the active site of the CYP109A2-HP-β-CD crystal structure. The black mesh represents a composite $2F_o$ - F_c omit electron density map, calculated at 1.77 Å resolution and contoured at 1 σ. C) Superposition of the three crystal structures determined in this study. The differences in the F and G helix conformations are depicted (2.7 Å substrate-free CYP109A2 structure (*pale yellow*), 2.2 Å CYP109A2-DMSO (chain A in *pink*; chain B in *green*) and 1.77 Å CYP109A2-HP-β-CD (*blue*)).

Overall structure of substrate-free CYP109A2

The crystal structure of substrate-free CYP109A2 was determined by the molecular replacement method to 2.7 Å resolution in an orthorhombic space group ($P\ 2_1\ 2_1\ 2_1$), with four monomers in the asymmetric unit and a solvent content of 50% (see Table 3 for data collection and refinement statistics). All four CYP109A2 molecules have well defined electron density, except for the N-terminal residues (including the accidental 26 amino acid N-terminal extension and following fifteen or sixteen residues) and the C-terminal His_6-tags, for which no electron density was observed. The central parts of the BC loop, FG loop and N-terminal end of the G helix show relatively high B factors, related to high mobility of those regions, as generally observed in P450 crystal structures. The four CYP109A2 molecules found in the asymmetric unit have almost identical conformations (overall Q score of 0.97; the root mean square deviation [r.m.s.d] of Cα-backbone atoms ranges between 0.36-0.43 Å). Therefore, only one of the protein molecules will be described further. The structure of CYP109A2 (residues 16-401) contains fifteen helices (labeled A-L, A', B' and K'), and ten β-strands (arranged in three β-sheets; in the same structural arrangement as observed in the CYP109E1 crystal

Table 3. Crystallographic data collection and refinement statistics of CYP109A2.

	Substrate-free CYP109A2	CYP109A2-DMSO[b]	CYP109A2-HP-β-CD[c]
Model statistics			
Monomers in the AU	4	2	1
Solvent content (%)	50	60	46
Ligands	5 x PEG 400 3 x Sulfate ion	n/a	Glycerol
Data collection			
Beamline (ESRF)	ID23-2	ID23-2	ID23-1
Wavelength (Å)	0.873	0.873	0.91165
Resolution range (Å)	52-2.7 (2.78-2.7)[a]	46-2.2 (2.26-2.2)	46-1.77 (1.81-1.77)
Space group	P 2_1 2_1 2_1	P 3_2 2 1	P 3_1 2 1
Unit-cell parameters			
a, b, c (Å)	77.12 155.53 158.15	78.47 78.47 325.05	78.37 78.37 128.40
α, β, γ, °	90 90 90	90 90 120	90 90 120
Observed reflections	287,529 (25,387)	452,310 (34,776)	593,422 (33,673)
Unique reflections	51,368 (4,458)	60,420 (4,425)	45,214 (2,548)
Multiplicity	5.6 (5.7)	7.5 (7.9)	13.1 (13.2)
Completeness (%)	97.3 (98.5)	99.9 (99.9)	100 (100)
$\langle I/\sigma(I) \rangle$	12.4 (1.7)	12.9 (2.9)	21.3 (1.7)
R_{merge} (%)	0.10 (0.94)	0.10 (0.61)	0.10 (1.70)
$R_{p.i.m.}$	0.07 (0.63)	0.05 (0.33)	0.04 (0.70)
CC (1/2) (%)	99.6 (52.3)	99.8 (86.2)	100 (63.9)
Refinement			
R_{work} (%)	22.38	19.77	19.19
R_{free} (%)	27.51	22.23	22.16
R.m.s.deviation, bond lengths (Å)	0.018	0.013	0.013
R.m.s.deviation, bond angles (°)	1.418	1.165	1.061
Average B-factors (Å)^2			
Overall	50.50	42.83	34.93
Protein	50.45	43.15	35.18
Heme	50.03	30.73	20.48
PEG 400, 1	54.36	n/a	n/a
PEG 400, 2	49.33	n/a	n/a
PEG 400, 3	56.21	n/a	n/a
PEG 400, 4	48.43	n/a	n/a
PEG 400, 5	74.76	n/a	n/a
SO4, 1	46.65	n/a	n/a
SO4, 2	75.81	n/a	n/a
SO4, 3	57.30	n/a	n/a
Glycerol (GOL)	n/a	n/a	37.96
Ramachandran plot statistics			
Most favoured (%)	95.6	96.6	97.1
Allowed regions (%)	4.4	3.4	2.9
Disallowed regions (%)	0	0	0
Molprobity overall score	1.49	1.38	1.17

[a] Values in parentheses refer to data in the highest resolution shells. [b and c] Crystal structures of CYP109A2 resulting from co-crystallization experiments involving VD_3 dissolved in DMSO and HP-β-CD, respectively.

structure, Figure 2A). The centrally located heme cofactor is sandwiched between the I and L helices, similarly as in other P450s. The bound heme is covalently attached to the protein by an iron-sulfur bond with Cys351, and is further stabilized by several residues providing van der Waals and hydrophobic interactions, while its propionate side chains interact with the side chains of His91, Arg95, Arg293 and His349. At the distal side of the heme, no clear electron density was observed for a heme-bound water molecule, in contrast to many other P450 crystal structures determined in the substrate-free state. A small blob of electron density is located at approximately 5 Å distance from the heme-iron, which may correspond to a glycerol molecule from the cryobuffer, but the quality of the electron density did not allow unambiguous ligand identification. In addition, five PEG 400 molecules and three sulfate ions (SO_4^{2-}), originating from the crystallization solution, are present in the structure.

Co-crystallization of CYP109A2 with vitamin D_3

To establish the binding mode of VD_3 we co-crystallized CYP109A2 with the compound (see Table 3 for relevant crystallographic statistics). Unfortunately, neither CYP109A2-DMSO or CYP109A2-HP-β-CD showed convincing density for a bound VD_3 molecule. In the CYP109A2-DMSO structure extra density was found in the active site that may represent a highly disordered VD_3 molecule, whereas in the CYP109A2-HP-β-CD structure the heme-iron is ligated by a glycerol molecule (relative occupancy of 80%) that originates from the crystal cryo-protecting solution and a partially bound water molecule (20%). The glycerol molecule is stabilized by a long hydrogen bond with the main chain carbonyl atom of Ala241 (SRS4). Two other water-mediated hydrogen bonds are present with (i) the main chain carbonyl oxygen atom of Ala290 (SRS5) and (ii) the side chain of Glu81 (SRS1). Glycerol binding is further supported by hydrophobic interactions with Thr245 (SRS4) and Ile288 (SRS5) (Figure 2B).

Both structures are in an open state and are highly similar to the CYP109A2 structure obtained in the absence of substrate (the r.m.s.d. ranges between 0.73-0.91 Å for Cα-backbone atoms). A small difference is observed in the F-G region, which shows a different degree of G helix bending in the CYP109A2-HP-β-CD and CYP109A2-DMSO structures (Figure 2C). The F-G region is mostly disordered in one of the CYP109A2-DMSO protein molecules (chain A).

Comparison of the structures of CYP109A2 with CYP109E1 and other related P450s

CYP109A2 crystallized in an ‘open’ conformation and its overall fold matches the canonical triangular-shaped P450 overall fold. A comparison with other P450 structures deposited in the PDB gives the best match to those displaying an open active site geometry and of which the enzymes have a similar substrate specificity as CYP109A2, such as CYP109E1, Vdh from *P. autotrophica*, and CYP109B1 from *B. subtilis* (Table 4).

The highest structural similarity is observed with the crystal structure of CYP109E1 in the open conformation (PDB entry 5L91, r.m.s.d. of 1.27 Å for 368 Cα atoms). The main differences are the presence of an additional α-helix in the N-terminal region of CYP109A2 (A’ helix, residues 24-28), and a differently folded BC loop (residues 68-88), which is significantly shorter and lacks the second 3_{10}-helix (Figure 3A). In addition, the G helix in CYP109A2 is displaced by ~4 Å towards the active site. Like in CYP109E1, the distal heme pocket in CYP109A2 is funnel-shaped, wide open and solvent exposed. Comparing the amino acid sequences of their common P450 substrate-recognition sites, the SRS4 regions of CYP109A2 and CYP109E1 show the highest similarity (78% identity)

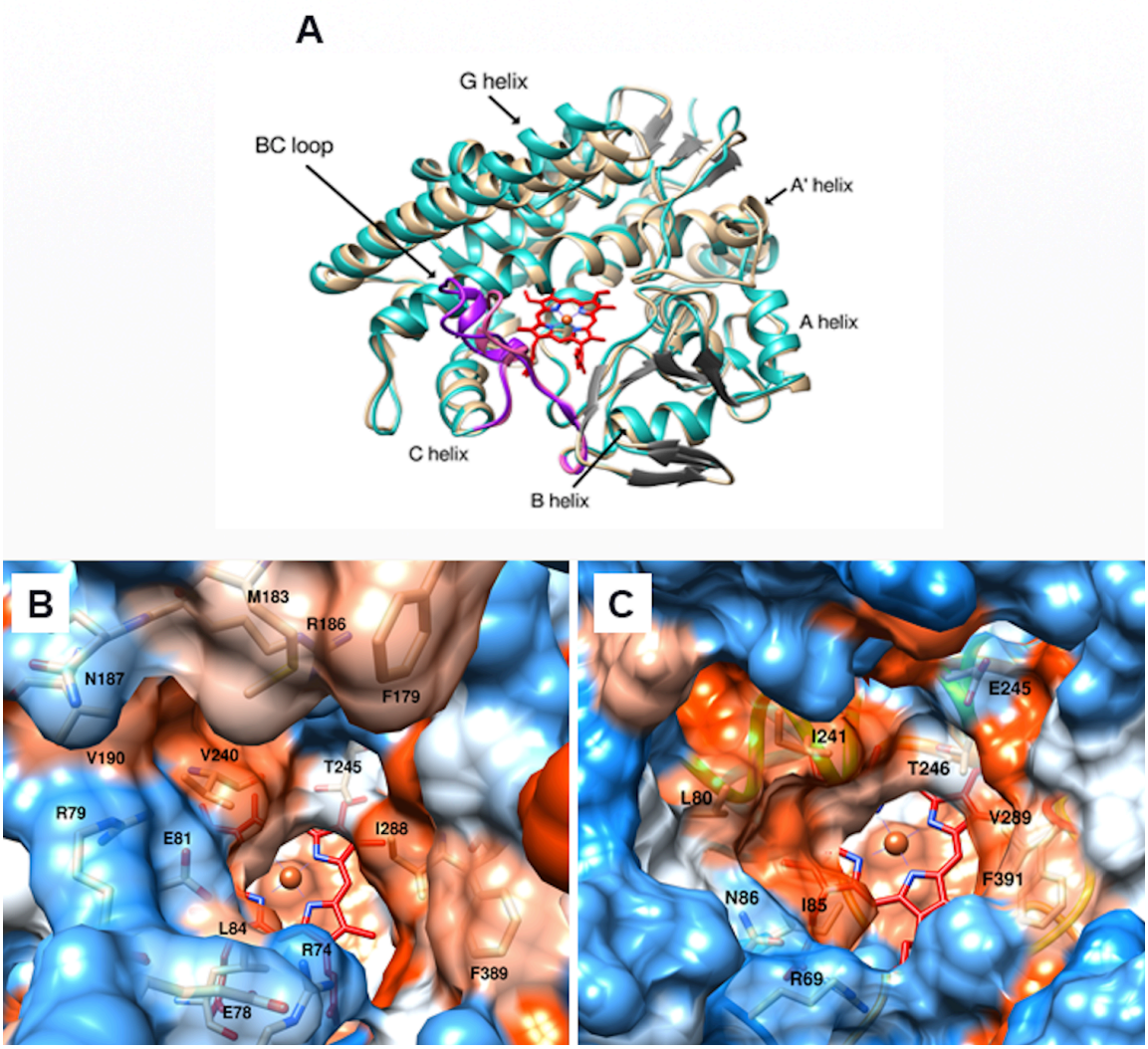

Figure 3. Comparison of the overall conformations and active sites of CYP109A2 and CYP109E1 from *B. megaterium* DSM319. A) Superposition of the substrate-free structures of CYP109A2 (*pale yellow*) and CYP109E1 (5L90, *cyan*). The heme is shown as *red* stick model. The BC loop is highlighted in *purple* in CYP109E1 and in *pink* in CYP109A2. Beta strands are in *grey*. B) and C) The distal pocket of the heme in the substrate-free CYP109A2 structure [B] versus the substrate-free structure of CYP109E1 [C]. Apolar residues are colored *orange,* polar residues are in *blue*.

related to the high conservation of the I helix residues in all P450s, followed by SRS5 (60% identity), SRS6 (60%) and SRS2 (F helix, 42%). The regions with the lowest sequence homology include SRS1 (BC loop, 21% identity) and SRS3 (G helix, no common residues). The structural differences of the BC loop are related to a striking difference in the distribution of hydrophobic and polar regions in the active sites of CYP109A2 and CYP109E1. Like in CYP109E1, the active site region in CYP109A2 that is closest to the heme (the 1st and 2nd ring of residues, Figure 3B and 3C) contains mainly hydrophobic residues: Leu84, Met85 (SRS1); Leu237, Val240, Ala241 (SRS4); Ile288, Ile291 (SRS5) and Phe389, Val390 (SRS6). Unlike CYP109E1, however, the wall of distal pocket of CYP109A2, formed by residues from the BC loop, is highly polar (residues Arg74, Glu78, Arg79 and Glu81). The BC loop in CYP109A2 also restricts accessibility to the heme-iron to a larger extent than the BC loop in CYP109E1 (Figure 3B).

Table 4. Highest ranked structural homologues of CYP109A2 and CYP109E1 found in the PDB. Three structures were subjected to a search with PDBeFold: CYP109A2 (open, substrate-free structure) and CYP109E1 (PDB id: 5L90 and 5L94, open and closed conformations, respectively).

CYP109A2 (open, substrate-free, chain C)						
	PDB code and chain id	Q score	RMSD Cα [Å]	N_{align}	% Sequence	P450 and organism
1	5L91:B	0.77	1.27	368	46	CYP109E1 *B. megaterium* DSM319
2	4RM4:A	0.67	1.44	339	47	CYP109B1 *B. subtilis*
3	4YZR:A	0.65	1.58	354	37	PksS *B. subtilis*
4	**3A4G:A**	**0.65**	**1.44**	**349**	**35**	**Vdh (CYP107) *P. autotrophica* (open)**
5	2BVJ:B	0.64	1.36	340	37	PikC (CYP107L1) *S. venezuelae*
6	2WIO:A	0.63	1.41	342	35	EryK (CYP113A1) *S. erythrea*
7	3EJE:D	0.62	1.85	357	35	BioI (CYP107H1) *B. subtilis*
8	5IKI:A	0.62	1.81	352	38	CYP106A2 *B. megaterium*
9	2YGX:D	0.61	1.62	347	32	MycG *Micromonospora griseorubida*
10	4FB2:A	0.60	1.59	342	30	$P450_{cin}$ *Citrobacter braakii*
CYP109E1 (open, substrate-free, 5L90, chain A)						
	PDB code and chain id	Q score	RMSD Cα [Å]	N_{align}	% Sequence	P450 and organism
1	n/a	0.76	1.34	368	46	CYP109A2 *B. megaterium* DSM319
2	4RM4:A	0.74	1.19	347	51	CYP109B1 *B. subtilis*
3	4YZR:A	0.69	1.48	359	35	PksS *B. subtilis*
4	3EJB:F	0.66	1.59	350	34	BioI (CYP107H1) *B. subtilis*
5	**3A4G:A**	**0.65**	**1.42**	**347**	**37**	**Vdh (CYP107) *P. autotrophica* (open)**
6	2BVJ:B	0.64	1.29	338	38	PikC (CYP107L1) *S. venezuelae*
7	2XKR:A	0.64	1.51	349	33	CYP142 *M. tuberculosis*
8	4TX3:A	0.63	1.53	337	28	OxyB *Actinoplanes teichomyceticus*
9	2WM5:A	0.63	1.30	349	32	CYP124 *M. tuberculosis*
10	1GWI:A	0.62	1.61	352	29	CYP154C1 *S. coelicolor* A3(2)
CYP109E1-TES (closed, 5L94, chain A)						
	PDB code and chain id	Q score	RMSD Cα [Å]	N_{align}	% Sequence	P450 and organism
1	3EJB:B	0.70	1.40	353	34	BioI (CYP107H1) *B. subtilis*
2	2WI9:A	0.69	1.42	360	37	PikC (CYP107L1) *S. venezuelae*
3	**3A4Z:A**	**0.67**	**1.31**	**352**	**36**	**Vdh (CYP107) *P. autotrophica* (closed)**
4	3R9B:B	0.67	1.30	346	37	CYP164A2 *M. smegmatis*
5	5CJE:A	0.67	1.31	345	35	CYP107L2 *Streptomyces avermitilis*
6	1JIP:A	0.66	1.44	354	33	EryF (CYP107A1) *S. erythrea*
7	4LHT:A	0.64	1.61	355	30	$P450_{cin}$ *Citrobacter braakii*
8	2JJO:A	0.64	1.41	344	38	EryK (CYP113A1) *S. erythrea*
9	4MM0:B	0.64	1.47	347	34	SgvP *Streptomyces griseoviridis*
10	2WM4:A	0.62	1.60	360	33	CYP124 *M. tuberculosis*

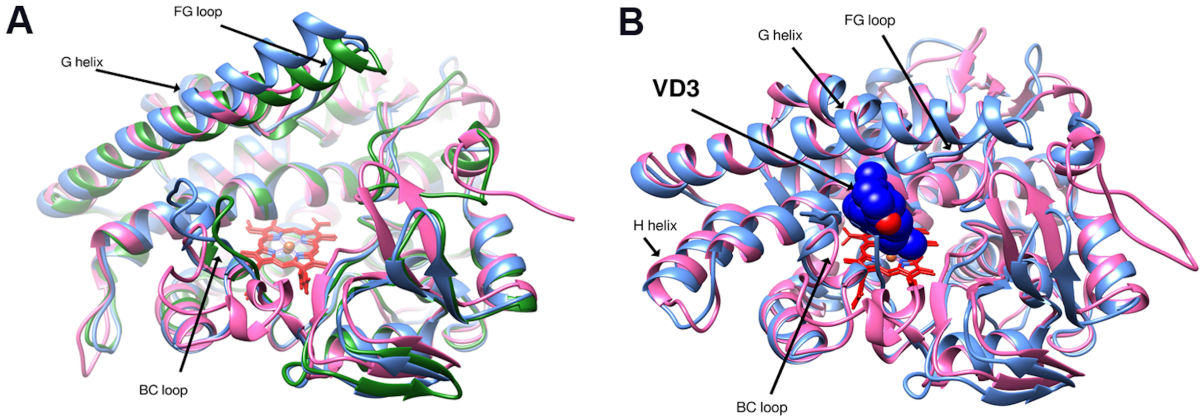

Figure 4. Superpositions of the structures of CYP109A2, CYP109E1 from *B. megaterium* DSM319 and Vdh from *P. autotrophica.* A) Cartoon representation of the aligned open conformations of substrate-free CYP109A2 (*green*), CYP109E1 (5L90, *blue*) and Vdh-WT (3A4G, *pink*). B) The closed conformations of CYP109E1 (5L94, *blue*) and Vdh-K1 (3A50, *pink*) are superimposed. The VD_3 molecule (*blue* spheres) is shown as bound in the crystal structure of Vdh-K1.

Besides its high similarity to CYP109E1, our comparison also reveals a significant similarity of CYP109A2 to Vdh from *P. autotrophica*. This similarity is interesting, since for the latter enzyme (Vdh-K1 mutant) crystal structures are available of both a substrate-free open form and a VD_3-bound closed form. While the crystallized open forms of either Vdh-WT or Vdh-K1 are similar to the open forms of CYP109A2 or CYP109E1, the VD_3-bound closed form of Vdh-K1 (PDB entry 3A50) matches very well the steroid-bound closed structure of CYP109E1 (r.m.s.d of 1.23 Å for 347 common Cα atoms, Figure 4), indicating that most probably the closed conformation of Vdh-K1 or CYP109E1 is the functionally relevant conformation of those P450s. Further research is needed to establish if the same holds true for CYP109A2.

Prediction of substrate-binding residues in CYP109A2

In order to understand why CYP109A2 is able to convert VD_3 to 25(OH)VD_3, but does not show any activity towards steroids, we compared the VD_3-binding residues in Vdh-K1 and the steroid binding residues in CYP109E1 to the equivalent residues in CYP109A2 (Table 5). The VD_3-binding residues of CYP109E1, as predicted by previous docking analysis [14], were also included in the comparison.

Comparing the residues that bind VD_3 in Vdh-K1 with the equivalent ones in CYP109A2 (Table 5) suggests that the following thirteen residues of CYP109A2 interact with VD_3: Arg74, Leu84, Met85 from SRS1; Leu167, Val168 from SRS2; Arg186, Asn187, Val190 from SRS3; Leu237, Val240, Thr245 from SRS4; Ile291 from SRS5 and Phe389 from SRS6 (Figure 5). Of these residues, six are identical to those in CYP109E1 or Vdh-K1, while the other residues are similar and can have similar interactions with the substrate.

Table 5. Identification of residues with predicted VD_3 and/or steroid binding roles in CYP109A2, based on structural comparison to CYP109E1 and Vdh-K1. A) Residues of CYP109E1 previously predicted to interact with the VD_3 molecule [14] compared to structurally equivalent residues in CYP109A2. B) Residues responsible for binding of VD_3 in the crystal structure of Vdh-K1 (PDB id: 3A50) [8] compared to structurally equivalent residues in CYP109A2. The residues that overlap with the ones predicted in panel A are in *blue*. C) Residues involved in steroid binding by CYP109E1 identified in the recently determined crystal structures [13], compared to structurally corresponding CYP109A2 residues. The residues of CYP109E1 studied by site-directed mutagenesis in our two previous works are highlighted in *grey*. Conserved CYP109A2 residues are shown in bold.

A	Predicted VD_3-binding residues CYP109E1	CYP109A2
SRS1	Arg69	**Arg74**
	Pro71	-
	Thr78	-
	Leu80	-
	Ile85	Leu84
	Asn86	Met85
SRS2	Ile168	Leu167
	Val169	**Val168**
SRS3	Asn191	Val190
SRS4	Leu238	**Leu237**
	Ile241	Val240
	Thr246	**Thr245**

B	VD_3-binding residues Vdh-K1 (3A50)	CYP109A2
SRS1	Thr84	-
	Met86	-
	Ile88	Leu84
	Leu89	Met85
SRS2	Leu171	**Leu167**
SRS3	Lys180	Arg186
	Asn181	**Asn187**
	Met184	Val190
SRS4	Leu232	**Leu237**
	Ile235	Val240
	Thr240	**Thr245**
SRS5	Pro287	Ile291
SRS6	Leu387	Phe389

C	Steroid-binding residues CYP109E1	CYP109A2
SRS1	Leu80	(Arg79/Glu81)
SRS2	Ile168	Leu167
	Val169	**Val168**
SRS4	Leu238	**Leu237**
	Ile241	Val240
	Ala242	**Ala241**
	Thr246	**Thr245**
SRS5	Val289	Ile288
	Ala291	**Ala290**
	His293	Gly292
SRS6	Phe391	**Phe389**
	Val392	**Val390**

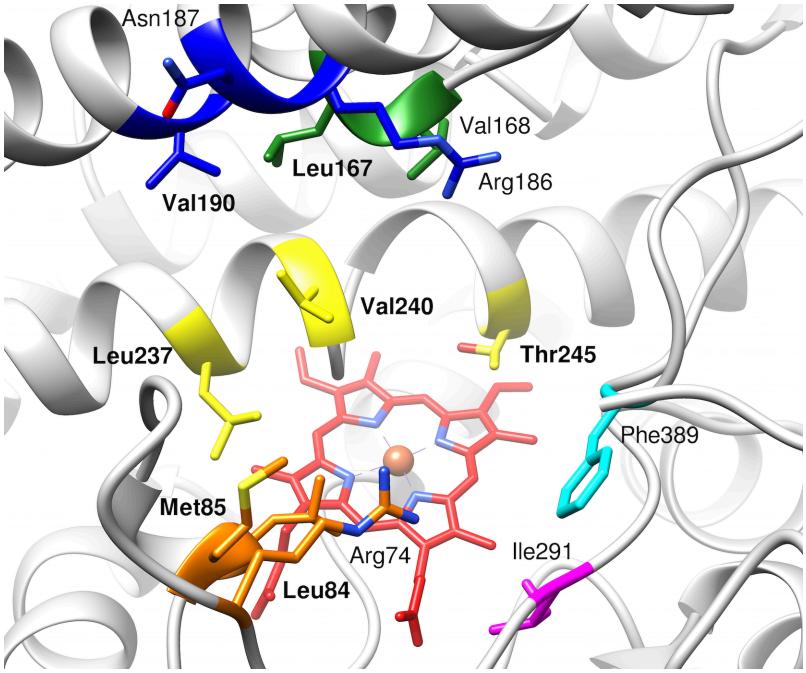

Figure 5. Putative VD_3-binding residues in CYP109A2. The active site residues predicted to bind VD_3 in CYP109A2, as based on a comparison with the CYP109E1 and Vdh-K1 structures (Table 5), are shown as sticks and coloured by SRS site (SRS1 *orange*, SRS2 *green*, SRS3 *blue*, SRS4 *yellow*, SRS5 *pink*, and SRS6 *cyan*). Residues identified in both predictions are labeled in bold. The heme group is shown in *red* coloured sticks.

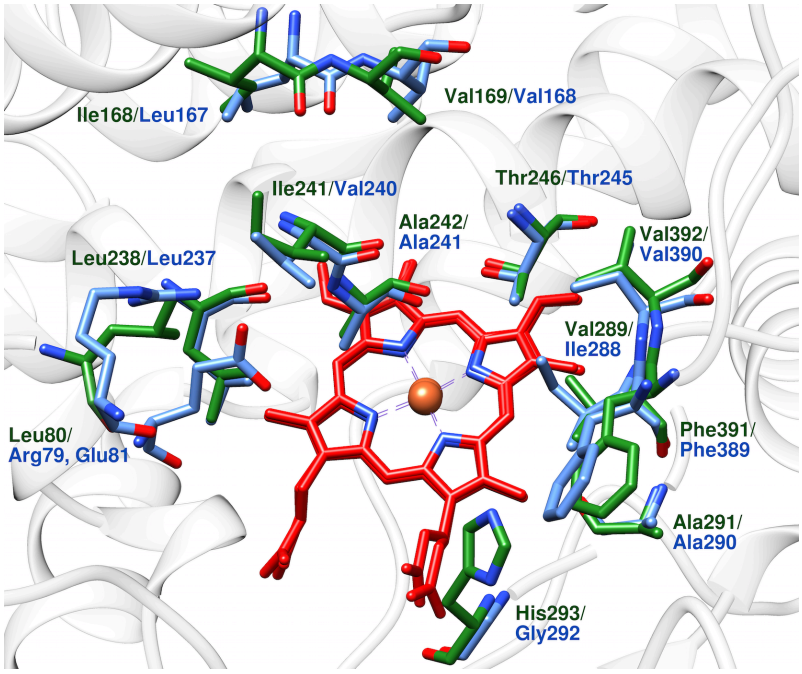

Figure 6. Overlay of important steroid binding residues in CYP109E1 (*green*) with their structural equivalents in CYP109A2 (*blue*). The residues predicted to have steroid-binding roles in CYP109A2 are hydrophobic residues generally very similar to those in CYP109E1. The most important difference concerns the BC loop residues: the hydrophobic residue Leu80 in CYP109E1 is replaced by Arg79 and Glu81 in CYP109A2, forming an active site pocket wall/region that could be unfavorable for steroid binding. The heme group is shown in *red* coloured sticks.

Our comparison further reveals that in CYP109A2 five out of twelve residues are different to those responsible for steroid-binding in CYP109E1 (Table 5). The most stricking variation is found in the BC loop region (SRS1). In CYP109E1, Leu80 in the BC loop interacts with steroids, but the CYP109A2 structure has no equivalent residue due to a different conformation of the BC loop. Instead, in CYP109A2 Arg79 and Glu81 are found at a similar location as Leu80 in CYP109E1, with their charged side chains facing the active site pocket (Figure 6). Since this part of the BC loop is not predicted to interact with VD_3, we conclude that these structural differences may relate to the absence of steroid binding activity in CYP109A2.

Discussion

CYP109A2 from *Bacillus megaterium* DSM319 is the first structurally characterized bacterial P450 that shows strict regio-selective hydroxylation activity at position 25 of VD_3. Other structurally characterized P450s that are able to hydroxylate VD_3, like CYP109E1 from the same organism and Vdh from *P. autotrophica*, show a significantly broader regio-selectivity. The crystal structures of CYP109A2 determined in this study all have an open conformation and are highly similar to structures of other CYP109 family members crystallized in their substrate-free states. The electron density was of sufficient quality to allow reliable modeling of residues 16 to 401 of CYP109A2, including the BC loop region (SRS1), which is a highly variable and flexible region in P450s and has been associated with differences in substrate specificity [32]. Unfortunately, co-crystallization experiments to obtain the structure of a VD_3-bound CYP109A2 were not successful. Possible reasons for failure are the low water solubility of VD_3 and the large flexibility of substrate and protein. Additionally, it may point to a low VD_3 binding affinity. Unfortunately, the dissociation constant of VD_3 could not be determined spectrophotometrically, due to the lack of the characteristic type I spectral change upon titration of VD_3 to CYP109A2 [12]. Also, it is not uncommon for P450s to crystallize in one state only, either open or closed, and to resist to be captured in the other state. CYP109B1, for example, was crystallized in its open form and all attempts to crystallize the protein with its substrates (β-ionone, oleic acid) were unsuccessful [33], producing only problematic crystals (low resolution, twinning).

To predict which residues have a role in VD_3 binding, the active site of CYP109A2 was compared to those of CYP109E1 and Vdh, two closely related bacterial P450s that also convert VD_3 to 25(OH)VD_3 and of which structures are available. Our results reveal that the residues in Vdh and CYP109E1 with a role in vitamin D_3 binding are significantly conserved in CYP109A2, indicative of a similar substrate binding mode. Comparison of CYP109A2 to other related VD_3-converting bacterial P450s further supports the identification of VD_3 binding residues in CYP109A2 (Figure 1). Interestingly, the two arginine residues Arg74 and Arg186 in CYP109A2, which are predicted to be involved in VD_3 binding, are equivalent to Arg73 and Arg193, respectively, in CYP105A1. The arginine residues in CYP105A1 are important for hydrogen bonding with the hydroxyl groups of VD_3 and its derivatives [34]. Whether the equivalent arginine residues in CYP109A2 have a similar role has to await further analysis by X-ray crystallography and site-directed mutagenesis.

No steroid converting activity could be observed for CYP109A2, which clearly differentiates this enzyme from CYP109E1. Our analysis suggests that sequence and conformation differences in the BC loop may be responsible for the different activities. While the BC loop in CYP109E1 contributes a hydrophobic residue to the active site pocket (Leu80), which participates in steroid binding, the BC loop residues at the same location in CYP109A2 are charged (Arg79/Glu81) and will likely prohibit steroid binding. Additionally, the BC loop in CYP109A2 restricts access to the active site pocket to a larger extent than in CYP109E1, which may prevent optimal access of the steroid molecule to the activated heme-iron. These findings are relevant for future research on CYP109A2 and may facilitate the choice of hotspots for protein engineering.

Materials and methods

Materials. Steroids tested in this work were bought from Sigma-Aldrich (Germany). Other chemicals were of highest purity available.

Expression plasmid. Plasmid pET17b_109A2, containing the *cyp109a2* gene from *B. megaterium* DSM319 (locus BMD_2035), was used for protein overexpression. The construction of this plasmid is described elsewhere [12]. The expressed protein construct contained 435 residues including: (i) residues 1-403 of wild-type CYP109A2, linked to (ii) an N-terminal extension of 26 amino acids (sequence MEKKSFIYSRLKTLPSFTFKRKKEDE), whose presence was due to an erroneous assignment of the starting codon, and (iii) a C-terminal His_6-tag.

Expression and purification. CYP109A2 was overexpressed in *E. coli* and purified using identical procedures (3-step purification) as previously described for CYP109E1 from *B. megaterium* DSM319 [13]. Typically about 16 mg of protein was obtained from 0.6 L culture. High protein purity and quality were confirmed by a single band in Coomassie stained 15% SDS-PAGE gels, by an absorbance ratio (A_{416nm}/A_{280nm}) of 1.9 obtained by spectrophotometric measurement and by a low percentage of polydispersity (12.2 %) obtained by dynamic light scattering. Purified protein was concentrated in 100 mM potassium phosphate buffer, pH 7.4, and 0.1 mM dithioerythritol (DTE), aliquoted and stored at -80 °C until further use.

Steroid binding assay. Spin-state shifts induced by the binding of steroids in the active site of CYP109A2 were analysed using a double-beam spectrophotometer (UV-2101PC, Shimadzu, Japan) and a pair of tandem quartz cuvettes, according to the method of Schenkman and Jansson [15]. In total, sixteen steroid compounds were tested: androstenedione (**1**), cortisol (**9**), cortisone (**10**), corticosterone (**2**), 11-deoxycortisol (RSS) (**11**), deoxycorticosterone (DOC) (**3**), dexamethasone (**12**), β-estradiol (**4**), estrone (**5**), 17α-methyltestosterone (**8**), 19-nortestosterone (**16**), prednisone (**13**), prednisolone (**14**), progesterone (**15**), testosterone (**6**) and testosterone acetate (**7**). One of the cuvette chambers was filled with 10 µM protein dissolved in 50 mM potassium phosphate buffer pH 7.4, while the other chamber contained buffer only. Steroids dissolved in dimethyl sulfoxide (DMSO) were added in increasing concentrations to the chambers (2.5 - 20 mM stock solutions) parallel to adding the corresponding amount of DMSO to the reference cuvette. Solutions were gently mixed and the spectral changes were recorded between 350 nm and 500 nm. All titrations were performed in triplicate.

***In vitro* conversion.** The *in vitro* conversions were performed for 30 min at 30 °C, in a final volume of 250 µl using 50 mM potassium phosphate buffer pH 7.4, containing 2% glycerol. The reaction mixture consisted of CYP109A2 (1 µM), a truncated form of bovine adrenodoxin (Adx_{4-108}, 20 µM), adrenodoxin reductase (AdR, 2 µM), an NADPH-regenerating system (1 mM $MgCl_2$, 5 mM glucose-6-phosphate, 1 U glucose-6-phosphate dehydrogenase), and 200 µM of the corresponding DMSO-dissolved steroid. The reactions were started by adding NADPH to a final concentration of 1 mM, stopped and extracted twice with 250 µl ethyl acetate. The samples were centrifuged (10,000 rpm, 10 min), organic phases were combined, dried by solvent evaporation, and, following resuspension, analyzed by high-performance liquid chromatography (HPLC).

HPLC analysis. HPLC analysis was done using a Jasco system (Pu-980 HPLC pump, AS-950 sampler, UV-975 UV/visible detector and LG-980-02 gradient unit; Jasco, Gross-Umstadt, Germany) and a reversed-phase ec MN Nucleodur C_{18} (3 µm, 4.0 x 125 mm) column (Macherey-Nagel, Bethlehem, PA, USA). The flow rate of the mobile phase was set to 1 ml/min and the system was kept at an oven temperature of 40 °C. The steroids were eluted by applying a gradient method, where the initial mobile phase ratio of 1:9 (acetonitrile:H_2O) was increased to 1:1. The UV detection of the compounds was accomplished at 240 or 254 nm.

Protein crystallization. Aliquots with purified CYP109A2 were thawed, buffer exchanged to 20 mM Tris/HCl pH 8.0, 0.1 mM DTE and concentrated to 40 mg/mL by ultrafiltration using a 30-kDa cutoff membrane. Crystallization condition screening was done in 96-well plate format. Sitting-drop vapor-diffusion experiments were set up by a Mosquito crystallization robot (TTP LabTech), using commercially available high-throughput crystallization screens (JCSG+, PACT and Structure, Molecular Dimensions; PEG/Ion, Hampton Research). Multiple hits were obtained overnight and the most promising condition (2.0 M ammonium sulfate, 0.1 M HEPES, pH 7.5, and 2% (v/v) PEG 400, *i.e.* Structure Screen condition C6) was chosen for further optimization. Crystal optimization was performed manually at 293 K by the sitting-drop vapour diffusion method in 24-well plates using reservoir solution volumes of 300 µL. A single, red plate-like crystal of CYP109A2 appeared after 3 days using a reservoir solution containing 1.9 M ammonium sulfate, 0.1 M HEPES, pH 7.5, and 2% (v/v) PEG 400 (protein drops were prepared by mixing 1 µL of protein and 1 µL of reservoir solution) and needed about a week to grow to its final dimensions (0.12 x 0.04 x 0.03 mm^3). Different co-crystallization experiments with VD_3 were set up, in an attempt to obtain crystals of the CYP109A2-VD_3 complex. The substrate was either dissolved in DMSO (100%) or 45% (w/v) 2-hydroxypropyl-β-cyclodextrin (HP-β-CD) in water, and added to diluted protein (~1.5 mg/mL) in 1:5 or 1:10 molar ratio. The mixtures were incubated for 1 h at 277 K, and then the protein was reconcentrated to 40 mg/mL, followed by crystallization screening as described above. Well diffracting crystals formed using the 1:5 mixture of CYP109A2 and DMSO-dissolved VD_3, with 0.1 M Bis-tris pH 5.5 and 1.7 M ammonium sulfate or the 1:10 mixture of CYP109A2 and VD_3 dissolved in 45% (w/v) 2-hydroxypropyl-β-cyclodextrin solution, with 0.1 M ammonium tartrate, pH 7.0, and 12% PEG 3350 (G5 condition of the PEG/Ion screen).

Data collection, structure determination and refinement. To protect the crystals against radiation damage during data collection a cryobuffer was prepared, consisting of mother liquor supplemented with glycerol (20% (v/v) for CYP109A2-HP-β-CD crystals and 30% (v/v) for substrate-free CYP109A2 and CYP109A2-DMSO crystals). The crystals were shortly swiped through a cryobuffer droplet and flash-cooled to 100 K in a cold nitrogen gas cryo stream. X-ray diffraction data were collected at the ID23-2 and ID23-1 beamlines of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. Reflections were indexed and integrated using XDS [16], while scaling and merging of the data was done with AIMLESS from the CCP4 software suite [17]. The initial structure of substrate-free CYP109A2 was determined by molecular replacement using Phaser from the PHENIX software package [18], applying the structure of CYP109E1 (46% sequence identity, PDB entry 5L90) as a search model. The structure was completed with several cycles of model refinement using phenix.refine [19] alternated with manual model re-building using Coot [20]. At the final stages of refinement and model building water molecules were added to the model based on positive difference electron density peaks and strict interaction criteria. The 2.7 Å crystal structure then served as a search model in molecular replacement to determine the other two crystal structures obtained by co-crystallization trials. Refinement of the CYP109A2-DMSO and CYP109A2-HP-β-CD structures proceeded similarly as the 2.7 Å CYP109A2 crystal structure.

Sequence and structure analysis. Protein sequences were compared using the Basic Local Alignment Search Tool (BLAST, NCBI). Multiple sequence alignment was done with Clustal Omega [21] and the output figure was made with ESPript3 [22]. The GenBank accession codes for analyzed sequences are as follows: CYP109A2 from *B. megaterium* DSM319 (WP_013082916), CYP109E1 from *B. megaterium* DSM319 (WP_013084555), CYP105A1 from *Streptomyces griseolus* (BAG50411), CYP105A2 from *Pseudonocardia autotrophica* (BAA05541), Vdh (CYP107BR1) from *P. autotrophica* **(**C4B644), CYP107CB2 from *B. lehensis* G1 (AIC83164) and human CYP2R1 (NP_078790.2). Substrate recognition sites (SRS) in CYP109A2 were identified by alignment with $P450_{cam}$ (WP_032492633) and followed a description defined by Gotoh [23]: SRS1: 71-87 (BC loop), SRS2: 163-169 (C' terminal part of F helix), SRS3: 185-192 (N' terminal part of G helix), SRS4: 229-247 (central part of I helix),

SRS5: 284-293 (K-β5 connection) and SRS6: 384-392 (β9-β10 turn). Pairwise structural alignments were done with PDBeFold [24]. The ligand binding interactions were analyzed with LigPlot+ [25].

Acknowledgements

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's 7th Framework Programme (FP7/2007-2013) under REA Grant Agreement 289217 (P4FIFTY). Authors would like to thank the beam-line scientists of ID23-1 and ID23-2 (ESRF, European Synchrotron Radiation Facility) for assistance during data collection.

Author contributions

IKJ, FMK, RB and AMWHT designed the study. IKJ, FMK, EB performed the experiments. IKJ, FMK, AMWHT analyzed the data. IKJ drafted the manuscript. FMK, BWD, RB and AMWHT assisted in drafting the manuscript. All authors read and approved the final manuscript.

References

1. Bernhardt R (2006) Cytochromes P450 as versatile biocatalysts. *J Biotechnol* **124**, 128-145.
2. Girvan HM & Munro AW (2016) Applications of microbial cytochrome P450 enzymes in biotechnology and synthetic biology. *Curr Opin Chem Biol* **31**, 136-145.
3. Bernhardt R & Urlacher VB (2014) Cytochromes P450 as promising catalysts for biotechnological application: chances and limitations. *Appl Microbiol Biotechnol* **98**, 6185-6203.
4. Donova MV & Egorova OV (2012) Microbial steroid transformations: current state and prospects. *Appl Microbiol Biotechnol* **94**, 1423-1447.
5. Zhu G & Okamura WH (1995) Synthesis of Vitamin D (Calciferol). *Chem Rev* **95**, 1877-1952.
6. Kametani T & Furuyama H (1987) Synthesis of vitamin D_3 and related compounds. *Med Res Rev* **7**, 147-171.
7. Sakaki T, Sugimoto H, Hayashi K, Yasuda K, Munetsuna E, Kamakura M, Ikushiro S & Shiro Y (2011) Bioconversion of vitamin D to its active form by bacterial or mammalian cytochrome P450. *Biochim Biophys Acta* **1814**, 249-256.
8. Yasutake Y, Fujii Y, Nishioka T, Cheon WK, Arisawa A & Tamura T (2010) Structural evidence for enhancement of sequential vitamin D_3 hydroxylation activities by directed evolution of cytochrome P450 vitamin D_3 hydroxylase. *J Biol Chem* **285**, 31193-31201.
9. Yasutake Y, Nishioka T, Imoto N & Tamura T (2013) A single mutation at the ferredoxin binding site of P450 Vdh enables efficient biocatalytic production of 25-hydroxyvitamin D(3). *Chembiochem* **14**, 2284-2291.

10. Hayashi K, Sugimoto H, Shinkyo R, Yamada M, Ikeda S, Ikushiro S, Kamakura M, Shiro Y & Sakaki T (2008) Structure-based design of a highly active vitamin D hydroxylase from *Streptomyces griseolus* CYP105A1. *Biochemistry* **47**, 11964-11972.

11. Hayashi K, Yasuda K, Sugimoto H, Ikushiro S, Kamakura M, Kittaka A, Horst RL, Chen TC, Ohta M, Shiro Y & Sakaki T (2010) Three-step hydroxylation of vitamin D_3 by a genetically engineered CYP105A1: enzymes and catalysis. *FEBS J* **277**, 3999-4009.

12. Brill E (2013) Identifizierung und Charakterisierung neuer Cytochrom P450 Systeme aus *Bacillus megaterium* DSM319, Universität des Saarlandes, URN: urn:nbn:de:bsz:291-scidok-55990.

13. Jóźwik IK, Kiss FM, Gricman L, Abdulmughni A, Brill E, Zapp J, Pleiss J, Bernhardt R & Thunnissen AW (2016) Structural basis of steroid binding and oxidation by the cytochrome P450 CYP109E1 from *Bacillus megaterium*. *FEBS J* **283**, 4128-4148.

14. Abdulmughni A, Jóźwik IK, Putkaradze N, Brill E, Zapp J, Thunnissen AW, Hannemann F & Bernhardt R (2016) Characterization of cytochrome P450 CYP109E1 from *Bacillus megaterium* as a novel vitamin D_3 hydroxylase. *J Biotechnol* **243**, 38-47.

15. Schenkman JB & Jansson I (1998) Spectral analyses of cytochromes P450. *Methods Mol Biol* **107**, 25-33.

16. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132.

17. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A & Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242.

18. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC & Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221.

19. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH & Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-367.

20. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132.

21. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD & Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539.

22. Robert X & Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-W324.

23. Gotoh O (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. *J Biol Chem* **267**, 83-90.

24. Krissinel E & Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256-2268.

25. Laskowski RA & Swindells MB (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model* **51**, 2778-2786.

26. Brill E, Hannemann F, Zapp J, Bruning G, Jauch J & Bernhardt R (2014) A new cytochrome P450 system from *Bacillus megaterium* DSM319 for the hydroxylation of 11-keto-beta-boswellic acid (KBA). *Appl Microbiol Biotechnol* **98**, 1701-1717.

27. Poulos TL, Finzel BC, Gunsalus IC, Wagner GC & Kraut J (1985) The 2.6-A crystal structure of *Pseudomonas putida* cytochrome P-450. *J Biol Chem* **260**, 16122-16130.

28. Hamdane D, Zhang H & Hollenberg P (2008) Oxygen activation by cytochrome P450 monooxygenase. *Photosynth Res* **98**, 657-666.

29. Kiss FM, Khatri Y, Zapp J & Bernhardt R (2015) Identification of new substrates for the CYP106A1-mediated 11-oxidation and investigation of the reaction mechanism. *FEBS Lett* **589**, 2320-2326.

30. Lee GY, Kim DH, Kim D, Ahn T & Yun CH (2015) Functional characterization of steroid hydroxylase CYP106A1 derived from *Bacillus megaterium*. *Arch Pharm Res* **38**, 98-107.

31. Schmitz D, Zapp J & Bernhardt R (2014) Steroid conversion with CYP106A2 - production of pharmaceutically interesting DHEA metabolites. *Microb Cell Fact* **13**, 81-2859-13-81.

32. Lepesheva GI, Virus C & Waterman MR (2003) Conservation in the CYP51 family. Role of the B' helix/BC loop and helices F and G in enzymatic function. *Biochemistry* **42**, 9091-9101.

33. Zhang A, Zhang T, Hall EA, Hutchinson S, Cryle MJ, Wong LL, Zhou W & Bell SG (2015) The crystal structure of the versatile cytochrome P450 enzyme CYP109B1 from *Bacillus subtilis*. *Mol Biosyst* **3**, 869-881.

34. Sugimoto H, Shinkyo R, Hayashi K, Yoneda S, Yamada M, Kamakura M, Ikushiro S, Shiro Y & Sakaki T (2008) Crystal structure of CYP105A1 ($P450_{SU-1}$) in complex with 1alpha,25-dihydroxyvitamin D3. *Biochemistry* **47**, 4017-402.